

Example on Aggregation

Part 1 _ Seppo Laaksonen

Micro data can be analysed ordinarily using statistical tools, including frequencies, averages, medians, distributional statistics, and more demanding methods as statistical models. The outputs of such techniques can be more or less as such published if they are useful. This kind of outputs are automatically aggregates. It is often good to save such outputs. It can be made manually, so that you type most interesting results in another file such as EXCEL, but this not always very nice to do. It is also possible to try to find more automatic tools. All software packages give opportunity to this. Next I present examples of ESS 5 data with SAS and SPSS.

My example is concerned income that has been based as everything on interviewing so that the result is 10 categories that are deciles of income. Unfortunately, it is possible to get income for everyone.

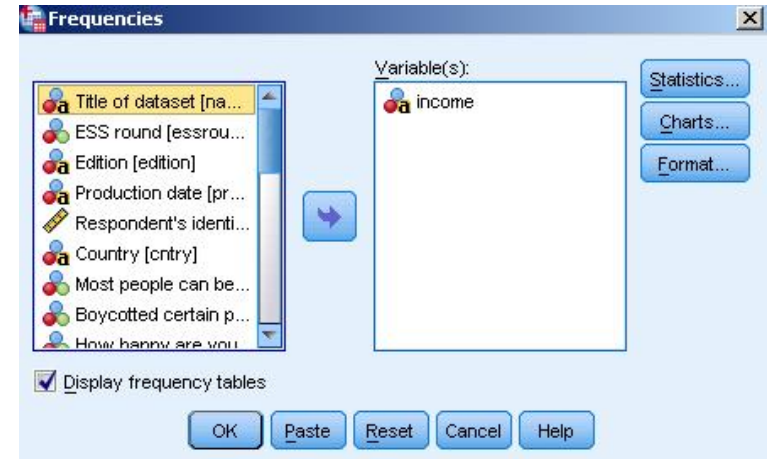
Aggregation example

The below table for all countries. I have reduced income categories as you see. There are three missing categories left.

income	Frequency	Percent	Cumulative Frequency	Cumulative Percent
10th decile	3245.195	6.19	3245.195	6.19
1st decile	3694.445	7.04	6939.641	13.23
2th-3rd deci	8191.49	15.61	15131.13	28.84
4th-5th deci	8555.808	16.31	23686.94	45.15
6th-7th deci	8273.665	15.77	31960.6	60.92
8th-9th deci	7055.147	13.45	39015.75	74.37
Missing	2150	4.10	41165.75	78.47
Other missin	4887.822	9.32	46053.57	87.79
Refusal	6406.103	12.21	52459.68	100.00

```
proc freq ;tables income; weight dweight; run;
```

SPSS OUTPUT HERE



income

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
	10th decile	3245	6,2	6,2
	1st decile	3694	7,0	13,2
	2th-3rd deci	8191	15,6	28,8
	4th-5th deci	8556	16,3	45,2
	6th-7th deci	8274	15,8	60,9
	8th-9th deci	7055	13,4	74,4
	Missing	2150	4,1	78,5
	Other missin	4888	9,3	87,8
	Refusal	6406	12,2	100,0
	Total	52460	100,0	100,0

Aggregation example

It is not nice that there are so much missingness. What do you think about Finland?

income	Frequency	Percent	Cumulative Frequency	Cumulative Percent
10th decile	133	7.08	133	7.08
1st decile	169	9.00	302	16.08
2th-3rd deci	312	16.61	614	32.69
4th-5th deci	316	16.83	930	49.52
6th-7th deci	432	23.00	1362	72.52
8th-9th deci	359	19.12	1721	91.64
Other missin	129	6.87	1850	98.51
Refusal	28	1.49	1878	100.00

Aggregation example

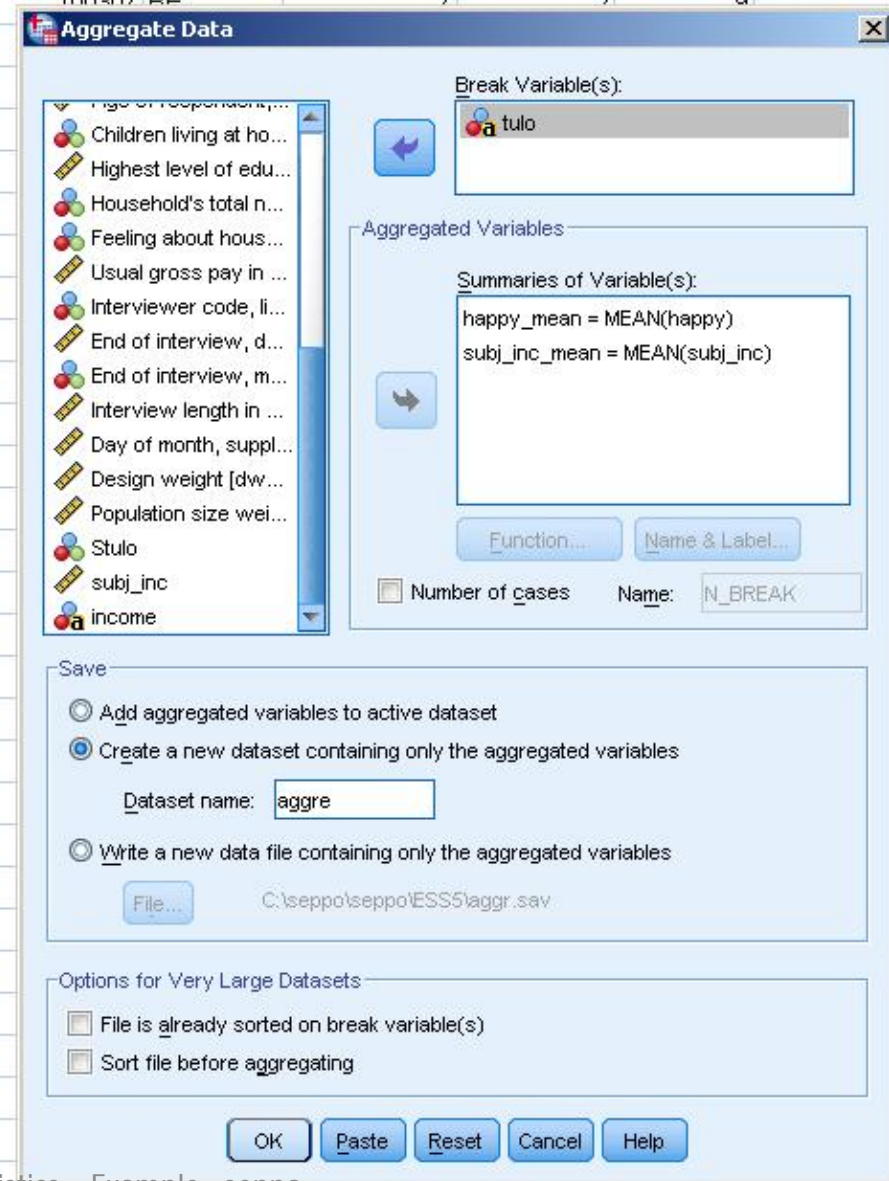
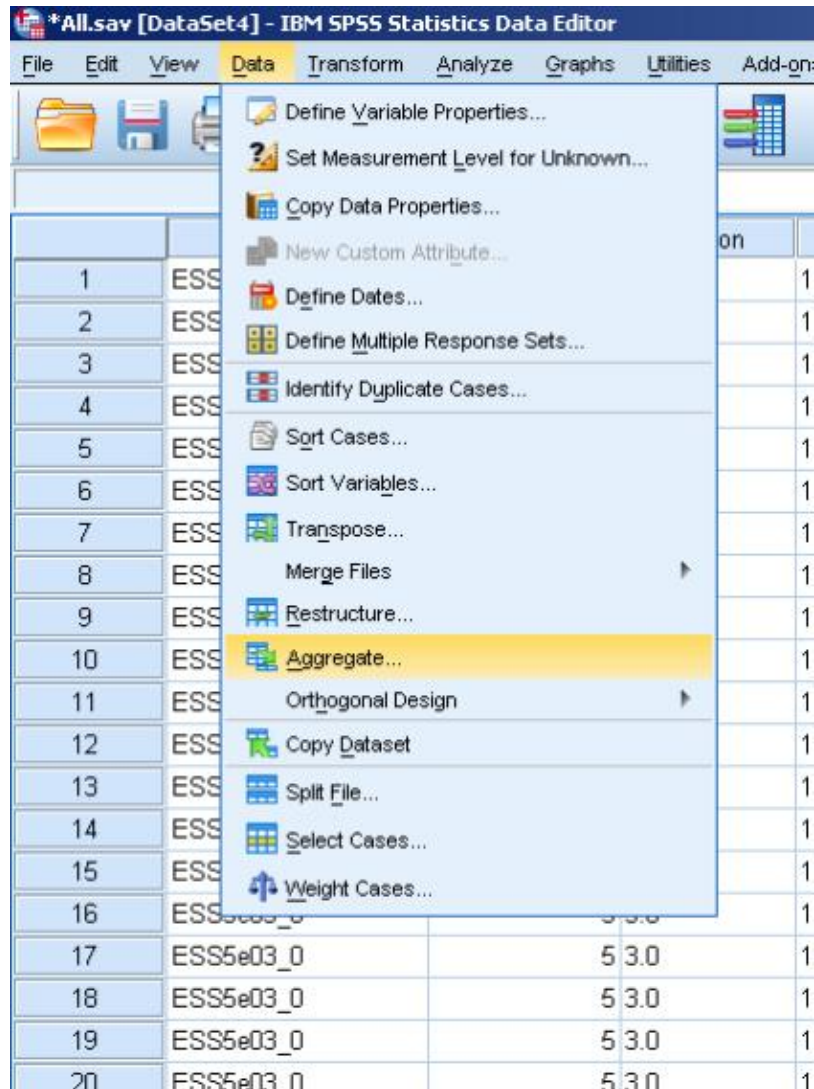
It would be nice to get some understanding what are those missing categories. One strategy is to look forward for variables without missingness or low missingness. I give an example in which two other variables are used:

- Happiness that varies from 0 (very unhappy) to 10 (very happy)
- Subjective income that varies initially from 1 (difficulties to live with such income) to 4 (very easy to live).

I have rescaled both so that they vary from 0 (very bad) to 100 (very good). It is good to note that both variables have some missingness, but much less than ordinary objective income. So, we can learn something.

Since I want to continue from basic results to graphical illustration, it is good to aggregate the output, i.e., to save it and continue then toward graphics.

Aggregation by SPSS



Aggregation by SAS

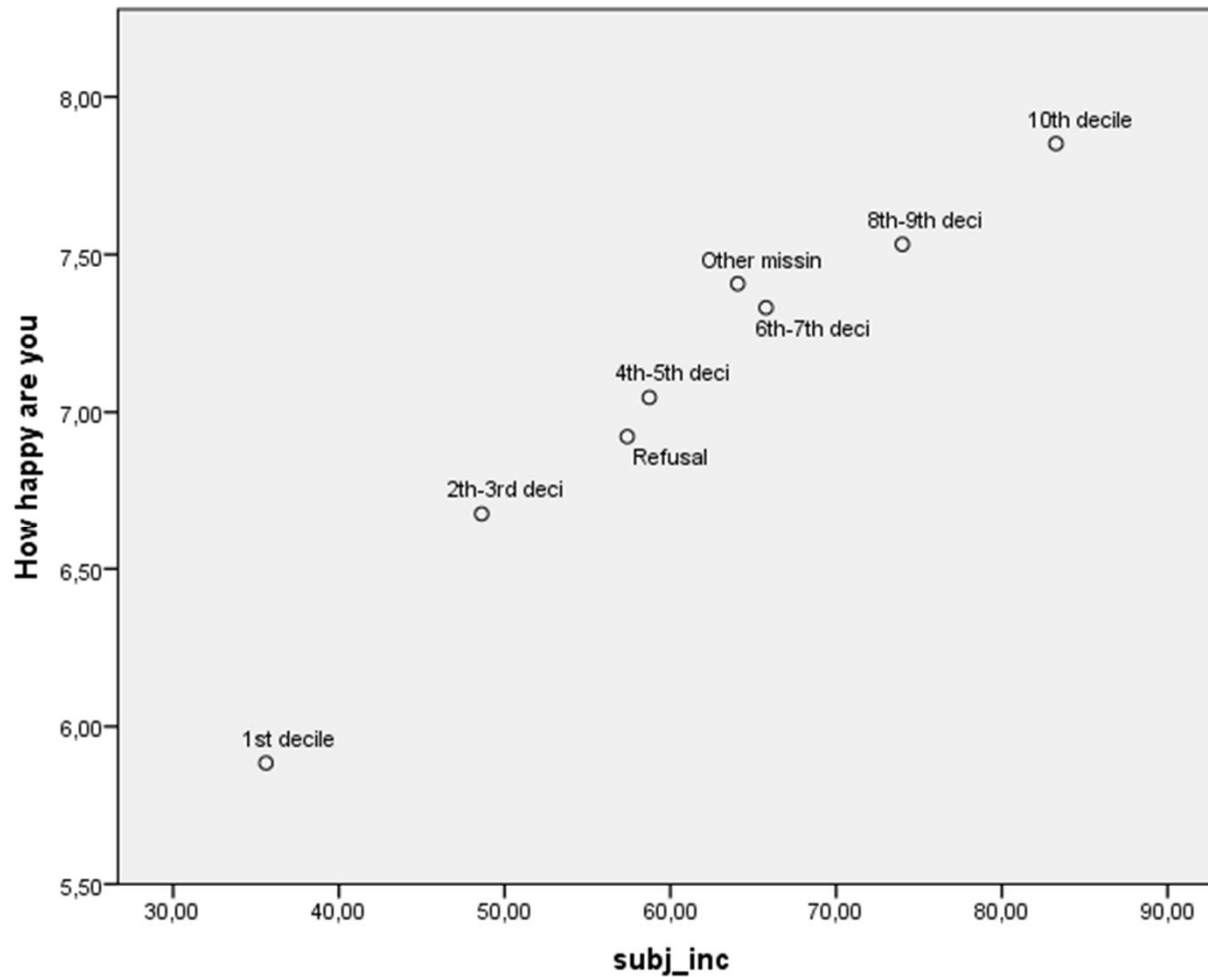
The weight is here as you see but SPSS weight should be selected before.

In both cases we will have an aggregate output file that can be further analysed as I promised. I will thus make a graph (scatter plot) but there can be more variables in the same file as well.

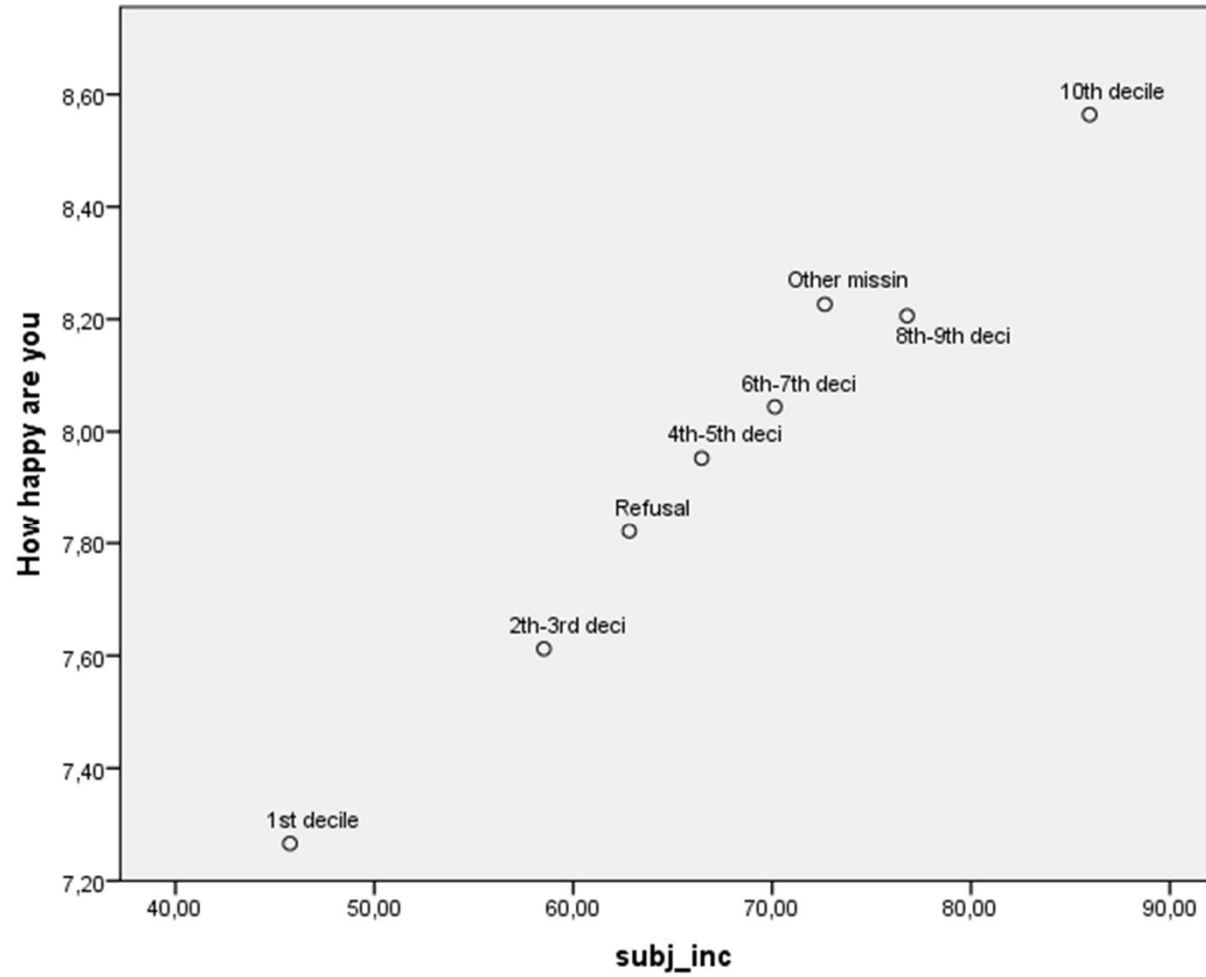
Try to interpret the next page graph for all countries! And one page later the same thing for Finland!

```
proc summary data=h.soc nway; *where cntry='FI'; class
income; var subj_inc happy ; output out=h.k_fi
mean(subj_inc happy)= subj_inc happy; weight dweight;
run;
```

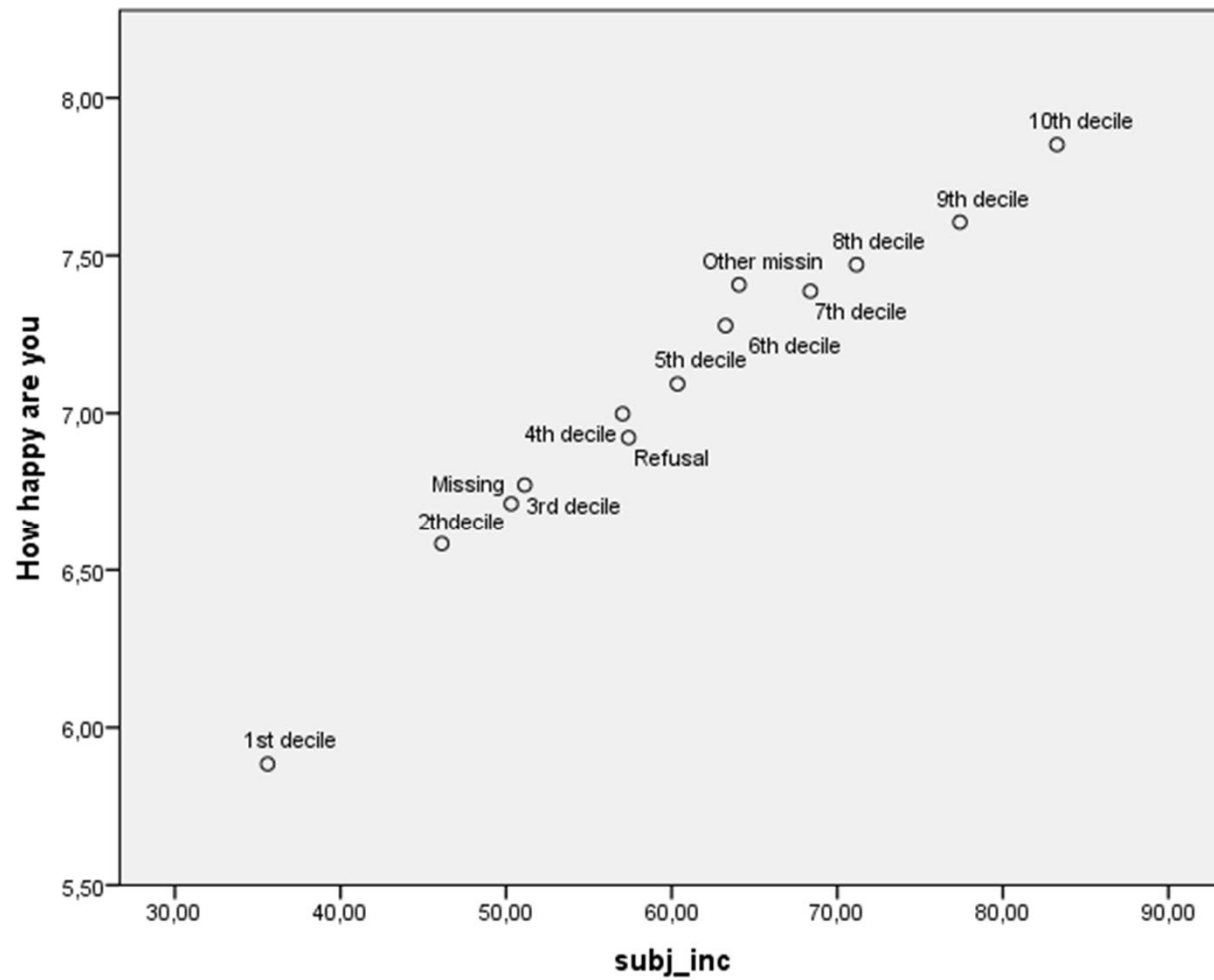
Scatter plot for all countries using SPSS Chart Builder
X-axis = Subjective income (average)
Y-axis = Happiness (average)



Scatter plot for Finland using SPSS Chart Builder



I made also a more detailed classification for income. What do you say about this for all countries?



What do you say about Finland's scatter ?

